

# Xanthusbase after five years expands to become Openmods

Philip C. Pratt-Szeliga<sup>1</sup>, Aaron D. Skewes<sup>2</sup>, Jinyuan Yan<sup>2</sup>, Laura G. Welch<sup>3</sup> and Roy D. Welch<sup>2,\*</sup>

<sup>1</sup>Department of Electrical Engineering and Computer Science, Syracuse University, Syracuse, NY, <sup>2</sup>Department of Biology, Syracuse University, Syracuse, NY and <sup>3</sup>Center for Advanced Systems and Engineering, Syracuse University, Syracuse, NY, USA

Received September 15, 2011; Revised October 11, 2011; Accepted October 25, 2011

## ABSTRACT

**Xanthusbase** (<http://www.xanthusbase.org>), a model organism database for the bacterium *Myxococcus xanthus*, functions as a collaborative information repository based on Wikipedia principles. It was created more than 5 years ago to serve as a cost-effective reference database for *M. xanthus* researchers, an education tool for undergraduate students to learn about genome annotation, and a means for the community of researchers to collaboratively improve their organism's annotation. We have achieved several goals and are seeking creative solutions to ongoing challenges. Along the way we have made several important improvements to Xanthusbase related to stability, security and usability. Most importantly, we have designed and implemented an installer that enables other microbial model organism communities to use it as a MOD. This version, called Openmods, has already been used to create Xenorhabdusbase (<http://xenorhabdusbase.bact.wisc.edu>), Caulobacterbase (<http://caulobacterbase.bsd.uchicago.edu>) and soon Bdellovibriobase.

## BACKGROUND: XANTHUSBASE AS A WIKI-STYLE COLLABORATIVE INFORMATION REPOSITORY

Wikis were first envisioned as collaborative information repositories (CIR) for the distributed annotation and re-annotation of genomes in model organism databases (MODs) more than 10 years ago (1). The idea held enormous promise. An online wiki-style CIR MOD offers the potential to solve three challenges confronted by many smaller model organism communities in the post-genomic era (2).

The first is the ongoing dependence of biological databases, including MODs, on constant federal support for

proper maintenance and curation (3). As databases proliferated over the past decade, many reached critical cross-roads: funding organizations that supported their design and implementation were anxious to wean them off the federal funding stream for maintenance and updates. Wikis were seen as one possible solution.

The second challenge is to develop an effective means to engage more undergraduate students in real genomics research. Undergraduates represent a large and enthusiastic group that can make meaningful advances in many areas of science (4), and genome annotation provides an excellent opportunity to realize this potential (5). This approach was first used to annotate the genome of the plant pathogen *Agrobacterium tumefaciens* C58 (6). Since then, there have been several additional annotation projects involving undergraduate students (5,7–9). Given these successes, it is somewhat surprising that the process has not been more widely adopted. The projects serve as a mechanism for undergraduate students to interact with researchers, contribute to genome annotations and see the impact of those contributions.

The third challenge faced by model organism communities is that, more often than not, researchers who study the model organism perceive that they have little or no direct participation in the annotation of its genome. Errors in the Genbank file for an organism's genome persist, sometimes for years, even though any member of that organism's research community could correct it with a few simple edits. Wikis were seen as a means of directly engaging a research community in the continual process of re-annotation, thereby offering a path to more accurate and up-to-date genomes (10).

Our original goal with Xanthusbase was to address each of these three problems. In 2005, at the 32nd International Conference on the Biology of the Myxobacteria, the research community that studies *Myxococcus xanthus* expressed unanimous support for the creation of an accessible, inexpensive wiki-style CIR MOD, and several Principle Investigators (PIs) also expressed interest in the educational potential of such a MOD. From its first

\*To whom correspondence should be addressed. Tel: +1 315 443 2159; Fax: +1 315 443 2012; Email: rowelch@syr.edu

iteration, the Xanthusbase user interface was designed to be familiar to anyone who has used a MOD. The genome annotation was parsed into genepages, each of which represented one open reading frame (ORF). Genepages were arranged in order along the genome, which was represented as a single circular chromosome. The feature that made Xanthusbase different from other databases was its editing feature. Any registered user could update the annotation by editing a genepage, with the following three caveats: all edits would be identified by author; all edits would be saved forever; and any user could return to a previous edit with a single mouse click. These three caveats were what made Xanthusbase 'wiki-style', since they are common to any wiki.

There are larger and more comprehensive databases that include information about the *M. xanthus* genome, such as NCBI (11), JCVI (<http://www.jcvi.org/>), BioCyc (12) and many others. Xanthusbase aggregates information from some of them, such as Pfam (13), KEGG (14) and GO (15), and reproduces some of the features of others, but its function as an autonomous community-owned and operated wiki-style CIR MOD remains unique. The continuous support of the community of researchers who study *M. xanthus* as a model organism, as well as interest from other small model organism communities, indicates that this system can play a unique role in supporting research and education.

## FIVE-YEAR UPDATE

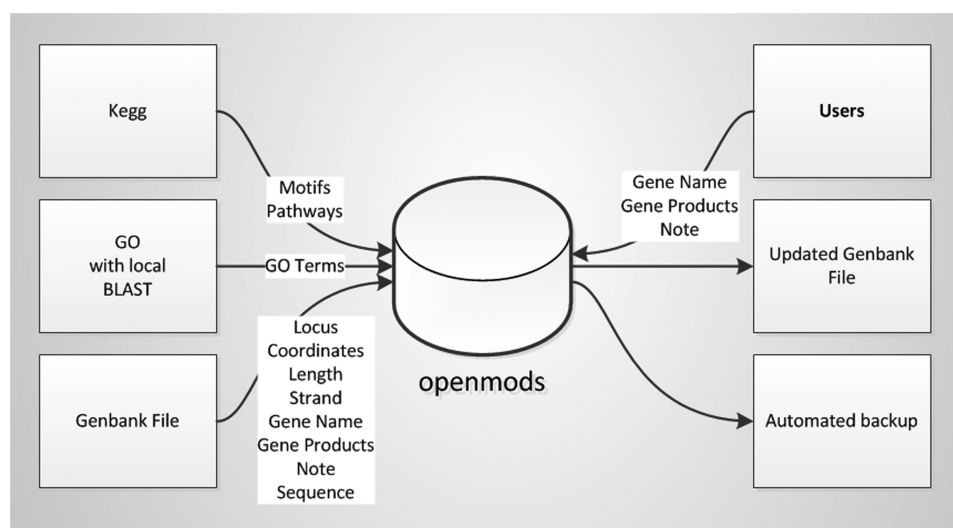
Since the original publication, we have made major improvements, many of which are not immediately obvious. We have completely rewritten the database software. The previous version required manual installation onto a server, but this made it difficult to create exact replica test servers for update development. Therefore, we

created an automatic installer to alleviate maintenance problems. The resulting installer and website software were general enough that it can be used to create a MOD for any bacterium that: (i) has a single circular chromosome, (ii) has an existing Genbank file and (iii) is listed in KEGG. Since the software is no longer specific to *M. xanthus* we renamed it Openmods; Xanthusbase now refers to the installation of Openmods for *M. xanthus*.

An Openmods installation requires a dedicated computer or virtual machine that runs Debian or Ubuntu and a dedicated network connection. The procedure is to execute two simple commands. The first installs the software dependencies that Openmods requires using the automated software installer present in Debian and Ubuntu (apt). The second command asks the user for some information (using debconf) about the specific installation. After about 10 min the website has acquired most of the information from external sources and is ready to be viewed. It takes about a week for the website to complete its first update of all information from external sources. Figure 1 gives a high level view of how information enters and exits the database.

Three steps are required from the user to begin an Openmods installation. The first is to set up the header; the user must provide a website name, PI information and institution logo. The top of Figure 2 shows the header of the Openmods installation specific to Xanthusbase. On the left hand side there is the text: 'XANTHUSBASE' with the subtitle: 'Myxococcus xanthus DK1622 and related bacteria'. This is a graphic that is automatically created by a Java program in the installer. On the right hand side there is PI contact information and the institution logo. After installation, this information can be updated using a command (\$openmods configure).

The next step is organism specification. The user is required to enter three pieces of information: the FTP location of the Genbank file, the KEGG organism



**Figure 1.** Openmods data sources and sinks. Motifs and Pathways are pulled using the KEGG SOAP API. Blast on Genes and Blast2GO are used to produce GO terms. The Genbank file is automatically downloaded via FTP, and the following information is then parsed from it: Locus, Coordinates, Length, Strand, Gene Name, Gene Products, Note and Sequence. Users are allowed to edit most of the fields, but Gene Name, Gene Products and Note are the only ones saved to the updated Genbank file. All data from the website is saved to the automatic backup.

# XANTHUSBASE

*Myxococcus xanthus* DK1622, and related bacteria

Roy Welch  
Department of Biology  
Syracuse University  
Syracuse, NY

[Home](#)
[Select Organism](#)
[Contact Us](#)
[Search](#)
[Update Status](#)

[Select Organism](#) > [Myxococcus xanthus](#) > [Gene Page](#)

**Genbank Information**

Locus: MXAN\_5767

Coordinates: 7142427 to 7143812

Length: 1386

Strand: Reversed

Gene Name: accC

Gene Name History		
Value	User	Date
accC	genbank	2009-11-11 00:00:00

**Gene Products:** acetyl-CoA carboxylase, biotin carboxylase

Gene Products History		
Value	User	Date
acetyl-CoA carboxylase, biotin carboxylase	genbank	2009-11-11 00:00:00

**Note:** identified by match to protein family HMM PF00289; match to protein family HMM PF02785; match to protein family HMM PF02786; match to protein family HMM TIGR00514

Note History		
Value	User	Date
identified by match to protein family HMM PF00289; match to protein family HMM PF02785; match to protein family HMM PF02786; match to protein family HMM TIGR00514	genbank	2009-11-11 00:00:00

**Motif Information**

Motif Id	Definition	Start	Stop	EValue	Date
pfam:CP_Sase_L_D2	Carbamoyl-phosphate synthase L chain, ATP binding domain	116	323	1.6e-77	2011-08-09 19:49:25
pfam:CP_Sase_L_chain	Carbamoyl-phosphate synthase L chain, N-terminal domain	1	109	7.1e-42	2011-08-09 19:49:25
pfam:Biotin_carb_C	Biotin carboxylase C-terminal domain	336	441	1.2e-37	2011-08-09 19:49:25
pfam:Dala_Dala_lig_C	D-ala D-ala ligase C-terminus	142	291	4.9e-10	2011-08-09 19:49:25
pfam:ATP-grasp	ATP-grasp domain	138	292	1.4e-08	2011-08-09 19:49:25

**Pathway Information**

Pathway	Description
mxm00061	Fatty acid biosynthesis - Myxococcus xanthus
mxm00062	Pyruvate metabolism - Myxococcus xanthus
mxm00064	Propanoate metabolism - Myxococcus xanthus
mxm01100	Metabolic pathways - Myxococcus xanthus
mxm01110	Biosynthesis of secondary metabolites - Myxococcus xanthus
mxm01120	Microbial metabolism in diverse environments - Myxococcus xanthus

**Gene Ontology Information**

Term 1	Relationship	Term 2
GO:0006259: DNA metabolic process	is_a	GO:0044260: cellular macromolecule metabolic process
GO:0006259: DNA metabolic process	is_a	GO:0090304: nucleic acid metabolic process
GO:0016421: CoA carboxylase activity	is_a	GO:0016885: ligase activity, forming carbon-carbon bonds
GO:0019842: vitamin binding	is_a	GO:0005488: binding
GO:0032559: adenylyl ribonucleotide binding	is_a	GO:0030554: adenylyl nucleotide binding
GO:0032559: adenylyl ribonucleotide binding	is_a	GO:0032555: purine ribonucleotide binding

**Mutant Strains**

Existing  
Mutant Strains: None

**Community Information**

**Protein Synonyms:**

Protein Synonyms History		
Value	User	Date
MXAN5767, Mx_4391, ORFA00006, ORFA6, GMXORF0036, GMXORF36, UniProtKB/TrEMBL:Q1D0B9	genbank	2009-11-11 00:00:00

**Gene Synonyms:**

Gene Synonyms History		
Value	User	Date
MXAN5767, Mx_4391, ORFA00006, ORFA6, GMXORF0036, GMXORF36, UniProtKB/TrEMBL:Q1D0B9	import	2009-11-14 00:00:00
MXAN5767, Mx_4391, ORFA00006, ORFA6, GMXORF0036, GMXORF36, UniProtKB/TrEMBL:Q1D0B9	genbank	2009-11-11 00:00:00

**Mutant Phenotypes:**

Mutant Phenotypes History		
Value	User	Date
	genbank	2009-11-11 00:00:00

**Community Description:** accB and accC genes are frequently located in a two-gene operon and regulated together to control biotin synthesis(delete)

Community Description History		
Value	User	Date
accB and accC genes are frequently located in a two-gene operon and regulated together to control biotin synthesis	Matthew Sejunt	2011-04-19 00:00:00
accB and accC genes are frequently located in a two-gene operon and regulated together to control biotin synthesis	genbank	2009-11-11 00:00:00

**Figure 2.** Openmods genepage screenshot. From top to bottom, the genepage contains: Genome browser, Locus (static), Coordinates (static), Gene Length (static), Strand (static), Gene Name (editable), Gene Products (editable), Note (editable), Motifs (periodically updated but not editable), Pathways (periodically updated but not editable), GO information (periodically updated but not editable), Mutant Strains (editable), Protein Synonyms (editable), Gene Synonyms (editable), Mutant Phenotypes (editable) and Community Description (editable). The genepage also contains the protein and nucleotide sequence but is not shown due to space constraints. The Gene Name, Gene Products and Note are included in Genbank file updates.

abbreviation and a name for the organism to be displayed in the 'Select Organism' dropdown menu. After the first organism is specified the installer continues onto the data backup section, but additional organisms can be installed on an Openmods installation using the command (`$openmods new_organism`). Organisms can also be deleted using (`$openmods delete_organism`)

The third step is for the user to specify how data backup will be performed. One of two options for automatic backup must be selected. The first option is to backup using a USB flash drive. If this option is selected, the user should insert a blank, 4GB (minimum) flash drive into any USB port on the computer. Once a week the automatic backup daemon will run and detect any flash drives plugged into the system and push backups there. Even if the entire computer hosting an Openmods installation is destroyed beyond repair, everything needed to recreate an installation is present on the flash drive, including usernames and user annotations. The procedure to install on a new computer from backup is: install a fresh copy of Openmods, do not specify an organism and run the command (`$openmods import <backup_filename>`). A backup can be triggered outside of the routine by running a command (`$openmods backup`). The second option is to send backups to a remote SFTP server. This is useful if Openmods is installed on a virtual machine where the administrator has no physical access, and therefore cannot plug in a flash drive. The administrator can change these settings at any time using the command (`$openmods backup_setup`).

Information is still organized according to genepages, where each genepage represents one ORF, and genepages are organized by their location in the genome. Figure 2 shows a representative genepage, where information can be grouped into sections from top to bottom. The first section is the Genome Browser, which is a clickable image showing nearby ORFs. A user can click an ORF in the Genome Browser and go to its genepage, or can scroll left or right along the chromosome by clicking on the left or right arrows. The Genome Browser was developed from scratch because existing genome browsers, specifically, GBrowse (<http://gmod.org/wiki/GBrowse>) require user intervention to install and do not integrate with the Openmods installer.

The second section contains data from the Genbank file. This section contains ORF Locus, Coordinates, Length, Strand, Gene Name, Gene Products and Note; Gene Name, Gene Products and Note are the only editable fields in this section. The third section contains Motif, Pathway and Gene Ontology data; none of this information is editable by the user. Motif and Pathway information is updated automatically from KEGG once a month, and the Gene Ontology information is updated using Blast2Go, also once a month. The fourth section contains fields that allow researchers to register mutant strains (suggested by the *M. xanthus* Community at the 38th Annual International Conference on the Biology of Myxobacteria). The fifth section contains information that is not in the Genbank file, but may be useful to the community, such as Protein Synonyms, Gene Synonyms, Mutant Phenotypes and Community Description.

Although none of the fields in this section are written out to an updated Genbank file, they are saved during backup.

It is not necessary to register in order to see Xanthusbase and browse the annotation; however, those who wish to add information must register. Access and security are managed by designating four types of users with different permission levels: blocked user, student, researcher and principle investigator. A person who registers with Xanthusbase will be a blocked user until their permission level is changed (to prevent web bots from destroying the integrity of the database). Users with permission level 'student' may contribute changes to the annotation, but these changes are marked as pending until approved by a user with permission level 'researcher' or 'principle investigator'. Users with permission level 'researcher' may make changes to the annotation without approval, and may approve or deny any suggested student-level changes. Users with permission level 'principle investigator' may do everything a researcher can do, and in addition can delete users, change their permission level, create groups of users according to laboratories or classes, and download a group's media (pdf's, images and videos) as a zip file with one folder per user in the group. This last feature is useful for grading. Also, only a 'researcher' or 'principle investigator' can register a mutant strain.

'Principle investigator' and 'researcher' contributions, as well as student proposed contributions, are visible for each ORF on its respective genepage, and they are also compiled on a list of recent changes. Although the entire history of changes is stored in the database in perpetuity, a user has the ability to immediately undo a change in the event of an input error. A Genbank file for the organism that incorporates all of the most recent changes can be downloaded using the command (`$openmods genbank_out`).

The discussion examines Openmods as a solution to each of the three problems stated in the introduction, using Xanthusbase as an example.

## ADDRESSING PROBLEM 1: OPENMODS AS AN FUNCTIONAL AND COST-EFFECTIVE MOD

Openmods has succeeded in becoming a functional and cost-effective MOD. It depends on only free software, is itself entirely free and open source, and can be installed on inexpensive hardware. It is built on PHP, Java, MySQL and the Joomla Framework. Joomla is used to allow administrators to change user's passwords and the static text on the site. For hardware, we recommend a computer with at least one 32-bit CPU core and 2 GB of RAM.

We have tested the current version of Openmods to ensure its reliability. Over the past year, Xanthusbase has been running 24 × 7 on the hardware configuration specified and received between 20 and 100 unique visitors per day. It can handle loads of at least 30 users simultaneously (all of the students in our class have simultaneously logged in and used the website). We currently have 74 registered users and over 7000 annotations.



Significant efforts have been made to ensure that the database remains secure. In all cases where content is uploaded according to a user level, Openmods conducts server-side checks for permissions. The site protects against SQL-injection attacks for all data that can touch an SQL query. It also protects against command line escaping when calling Java-based programs for backend processing. User content is filtered for JavaScript injection or cross-site scripting. User passwords are stored as MD5 hashes with salt by Joomla so that even people with administrator access cannot view a user's password.

## ADDRESSING PROBLEM 2: OPENMODS IN UNDERGRADUATE BIOINFORMATICS EDUCATION

For the past 5 years, we have taught an introductory bioinformatics class designed for upper-division undergraduate students with a strong background in molecular biology but limited computer skills. Its purpose is to present a subset of the fundamental file formats, algorithms and databases in sufficient detail that the students are able not only to identify and apply them, but also to make informed and critical interpretations. Our curriculum has lecture and computer laboratory components that teach genome annotation, and students must then apply this knowledge in interactive research. Each lecture provides background for the subsequent laboratory, which ultimately leads to a final project where students contribute to the re-annotation of a bacterial genome (usually, but not always *M. xanthus*). Lecture topics and the laboratory exercises for 2011 are included in [Supplementary Data](#).

We introduced a re-annotation exercise into the course using Xanthusbase in 2009. In 2011 we greatly expanded the role of Xanthusbase, due both to curriculum development and the aforementioned significant software redesign. For the final project, students were assigned approximately 50 ORFs from the *M. xanthus* genome and asked to perform a thorough and documented re-annotation using Xanthusbase.

The quality of the 2011 student annotations exceeded our expectations, thus improving the annotation for 136 ORF's when compared to the current *M. xanthus* Genbank file (<http://www.ncbi.nlm.nih.gov/nucore/108460647>). Approximately 65% of the contributions were approved as being meaningful, with the most productive student improving the annotation of all 50 reading frames she was assigned. All contributions remain on Xanthusbase and can be viewed by anyone (Select Organism / *Myxococcus xanthus* / Recent Annotations).

## ADDRESSING PROBLEM 3: OPENMODS AS A WIKI-STYLE CIR MOD FOR THE RESEARCH COMMUNITY

The expected area of impact where we have encountered the most challenges has been in engaging the *M. xanthus* research community in active, collaborative genome annotation through Xanthusbase. While certainly some

researchers and students have made meaningful contributions, there have been far fewer than was originally envisioned. We believe that this does not reflect negatively on Openmods, which has proven itself over the past 5 years to be a reliable MOD and an extremely useful teaching tool, and we certainly do not believe it reflects negatively on the community of *M. xanthus* researchers, who represent a truly dedicated and collaborative group. Wiki-style databases have proven to be a highly effective means of collecting and organizing information under certain circumstances, but they are not always the best option.

The barriers to collaborative genome annotation have been discussed, and these remain unchanged since life science wikis began to proliferate. Although there has been no rigorous examination as to why academic researchers do not embrace wiki-style CIRs, there are reasons that seem obvious to an inside observer. Academic research takes place within an established incentive structure defined by a series of goals: a tenure-track job, research funding, tenure and promotion. Contributions to a wiki-style CIR do not move a researcher closer to these goals. In fact, such contributions may actually function as a hindrance, since data contributed to a wiki-style CIR might be considered published and therefore ineligible for subsequent publication in a peer-reviewed journal. Even if that were not the case, explaining your ideas pre-publication in an online forum could be viewed as potentially helping your competitors. Hopefully, the academic incentive structure will evolve to incorporate new approaches. For example, Prof. Michel Aaij at Auburn University Montgomery recently used his more than 60 000 contributions to Wikipedia as part of his tenure portfolio (<http://blog.wikimedia.org/2011/04/06/tenure-awarded-based-in-part-on-wikipedia-contributions>), but at present he represents an extreme outlier.

Although PIs may not have an adequate incentive to contribute to a wiki-style CIR MOD, perhaps graduate students and post-doctoral fellows would contribute? Since they are already experts, it could be argued that contributing would require little additional effort. Unfortunately, in a study that analyzed the effect of wikis on student writing, Wheeler *et al.* (16) found that students who contributed to wikis were cognizant that their contributions lacked anonymity, and would be read by peers, and that this caused them to spend more time and write more carefully than usual for their assignments. The same is likely to apply to graduate students and post-doctoral fellows contributing to a wiki-style CIR MOD that might be read by their peers, PI and other members of the scientific community. Perhaps because of this, they would feel the need to devote considerable time and effort to their contributions, and thus may arrive at the conclusion that this time and effort are better spent writing a manuscript or generating more data.

Unlike graduate students, there is a strong incentive for undergraduate students to make contributions to a wiki-style CIR MOD: they want to earn a good grade in class. They are not yet experts in the model system or the process of annotation, however, and so their contributions vary in quality. This is apparent in the recent

contributions from our 2011 bioinformatics class; some provide real insight into the annotation of an ORF, but others add little information. Any database built on Wiki principles relies on user contributions to be accurate and thorough. If Xanthusbase is to succeed, we must continue working toward a sustainable community-driven model. In this way, Xanthusbase is an ongoing experiment in scientific social networking. Having expanded to other model organisms and research communities with Openmods, we will try to learn from each other over the next 5 years. We continue to look for new participants.

## CONCLUSIONS: OPENMODS 5-YEAR PLAN

Openmods software will be continually updated in the next 5 years. We plan to create Openmods installations for new model organisms, and to support organisms with multiple chromosomes and also chromosomes that are not circular. One PI (S. Crosson, personal communication) has requested that we allow changing the coordinates of ORFs and also removing and adding ORFs. We plan to add this functionality. We will also introduce a feature that allows users of an Openmods installation to vote on the importance of an annotation (similar to the voting in [stackoverflow.com](http://stackoverflow.com)), and show records of users by the number of votes they have accumulated. We plan to work with the Debian open source community to make the Openmods package available in the Debian repositories. This will make installation easier, and we will be able to push updates to Openmods installations automatically, allowing us to more easily implement new community feature requests and respond to security defects that require patches. We are also looking into using web frameworks that feature unit testing. This will allow us to more readily accept contributions from open source developers because it will allow us to ensure that contributions from unknown developers are of sufficiently high quality.

We also plan to expand the use of Openmods as an undergraduate education tool. We are seeking other PI's outside of our current set to incorporate Openmods into their curriculum. Moreover, we envision a collaborative environment among each model organism's group where the community will act as ad hoc referees for the undergraduate contributions. This will not only ensure that high-quality contributions become fixed in the database, but will teach necessary skills to the next generation of life scientists. The benefits of introducing substantive research projects into the classroom are real (17). Undergraduate involvement in research can improve retention in the sciences (18) and the pursuit of a graduate degree (19). In a 2004 study of 1135 undergraduates from 41 universities, 91% of students reported that their research experience sustained or increased their interest in attending graduate school (20). Based on our experience with the success and continued challenges of Openmods, it seems a natural progression to encourage more undergraduate participation in genome annotation.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online. Supplementary Lab 1–7.

## ACKNOWLEDGEMENTS

We would like to acknowledge the students who were in BIO422 (bioinformatics for life scientists) and contributed the most accepted re-annotations, especially Jiaye Guo (50 contributions), Stephanie Jones (47) and Matthew Sejuitt (21). We would also like to thank Prof. Heidi Goodrich Blair, Prof. Sean Crosson, and Prof. Mark Martin, who have agreed to host Xenorhabdusbase, Caulobacterbase, and Bdellovibriobase respectively.

## FUNDING

Funding for open access charge: National Science Foundation (MCB-0746066, to R.D.W.).

*Conflict of interest statement.* None declared.

## REFERENCES

- Dowell, R.D., Jokerst, R.M., Day, A., Eddy, S.R. and Stein, L. (2001) The distributed annotation system. *BMC Bioinformatics*, **2**, 7.
- Arshinoff, B.I., Suen, G., Just, E.M., Merchant, S.M., Kibbe, W.A., Chisholm, R.L. and Welch, R.D. (2007) Xanthusbase: adapting wikipedia principles to a model organism database. *Nucleic Acids Res.*, **35**, D422–D426.
- National Science Foundation (2005). *US NSF - NSB-05-40, Long-Lived Digital Data Collections. Enabling Research and Education in the 21st Century*. <http://www.nsf.gov/pubs/2005/nsb0540/start.htm>.
- Goodner, B.W., Wheeler, C.A., Hall, P.J. and Slater, S.C. (2003) Massively parallel undergraduates for bacterial genome finishing. *Asm. News*, **69**, 584–585.
- Forst, S. and Goodner, B. (2006) Comparative bacterial genomics and its use in undergraduate education. *Biol. Control*, **38**, 47–53.
- Slater, S., Goodner, B., Hinkle, G., Gattung, S., Miller, N., Blanchard, M., Qurollo, B., Goldman, B.S., Cao, Y.W., Askenazi, M. et al. (2001) Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. *Science*, **294**, 2323–2328.
- Banerjee, U., Chen, J., Call, G.B., Beyer, E., Bui, C., Cespedes, A., Chan, A., Chan, J., Chan, S., Chhabra, A. et al. (2005) Discovery-based science education: functional genomic dissection in *Drosophila* by undergraduate researchers. *PLOS Biol.*, **3**, 207–209.
- Setubal, J.C., Slater, S.C., Goldman, B.S., Goodner, B., Farrand, S.K., Nester, E.W., Burr, T.J., Banta, L., Dickerman, A.W., Paulsen, I. et al. (2009) Genome sequences of three *Agrobacterium* biovars help elucidate the evolution of multichromosome genomes in bacteria. *J. Bacteriol.*, **191**, 2501–2511.
- Elgin, S.C.R., Shaffer, C.D., Alvarez, C., Bailey, C., Barnard, D., Bhalla, S., Chandrasekaran, C., Chandrasekaran, V., Chung, H.M., Dorer, D.R. et al. (2010) The genomics education partnership: successful integration of research into laboratory classes at a diverse group of undergraduate institutions. *Cbe-Life Sci. Educ.*, **9**, 55–69.
- Salzberg, S.L. (2007) Genome re-annotation: a wiki solution? *Genome Biol.*, **8**.
- National Center for Biotechnology Information (U.S.). 2002 *NCBI*, Bethesda, MD.
- Caspi, R., Altman, T., Dale, J.M., Dreher, K., Fulcher, C.A., Gilham, F., Kaipa, P., Karthikeyan, A.S., Kothari, A., Krummenacker, M. et al. (2010) The MetaCyc database of

- metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.*, **38**, D473–D479.
13. Finn,R.D., Mistry,J., Tate,J., Coghill,P., Heger,A., Pollington,J.E., Gavin,O.L., Gunasekaran,P., Ceric,G., Forslund,K. *et al.* (2010) The Pfam protein families database. *Nucleic Acids Res.*, **38**, D211–D222.
14. Kanehisa,M. and Goto,S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*, **28**, 27–30.
15. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene Ontology: tool for the unification of biology. *Nat. Genet.*, **25**, 25–29.
16. Wheeler,S. and Wheeler,D. (2009) Using wikis to promote quality learning in teacher training. *Learn. Media Technol.*, **34**, 1–10.
17. Council,N.R. (2003) *BIO2010: Transforming Undergraduate Education for Future Research Biologists*. The National Academies Press, Washington, D.C., p. 192.
18. Taraban,R. and Blanton,R.L. (2008) *Creating Effective Undergraduate Research Programs in Science : the Transformation from Student to Scientist*. Teachers College Press, New York, NY.
19. Bauer,K.W. and Bennett,J.S. (2003) Alumni perceptions used to assess undergraduate research experience. *J. High Educ.*, **74**, 210–230.
20. Lopatto,D. (2004) Survey of Undergraduate Research Experiences (SURE): first findings. *Cell Biol. Educ.*, **3**, 270–277.